Robust 3D Face Recognition from Low Resolution Images

A.Drosou[1,2], P.Moschonas[1], D.Tzovaras[1].
[1]*Centre for Research and Technology Hellas, Information Technologies Institute*
57001, Thessaloniki, Greece
[2]*Imperial College London, Department of Electrical Engineering, London, UK*
{*drosou,moschona,tzovaras*}*@iti.gr*

Abstract: This paper proposes a combined approach for robust face recognition from low resolution images captured by a low-budget commercial depth camera. The low resolution of the facial region of interest is compensated via oversampling techniques and efficient trimming algorithms for the generation of an accurate $3D$ facial model. Two state of the art algorithms for geometric feature extraction are then utilized, i.e. the estimation of the Directional Indices between all the isogeodasic stripes of the same facial surface via the $3D$ Weighted Walkthroughs ($3DWW$) transformation and the estimation of the Spherical Face Representation (SFR). The biometric signature is then enhanced via user-specific cohort biometric templates for each feature, respectively. The experiments have been carried out on the demanding "BIOTAFTOTITA" dataset and the results are very promising even under difficult scenarios (e.g. looking away instances, grimace, etc.). Despite the obvious superiority of the $3DWW$ transformation over the SFR, it has been noted that the score level fusion of both algorithms improves the authentication performance of the system. On the contrary, only the $3DWW$ transformation should be preferred in identification scenarios. Indicatively, the experimental validation on the aforementioned dataset containing 54 subjects illustrates significant succeeds an identification performance of $\sim\ 100\%$ in Rank-1 and Equal Error Rate of 0.25% regarding the authentication performance in the neutral face experiment.

1 Introduction

It is a common place that security in computer systems is an increasingly critical issue that affects a series of diverse applications, ranging from granting access control in restricted infrastructures to e-commerce transactions. Such applications require reliable personal recognition schemes to either confirm or determine the ID of an individual requesting their services. To this extent, biometrics have been proven to provide unique and powerful advantages over other traditional technologies for ID verification (e.g. PINs, tokens, etc.) that can be easily forgotten, lost or stolen.

Human recognition systems have long been developed based on biometric characteristics of the person, such as in [Ross2003a] and in [Tsalakanidou07]. Although, researchers have long been working on a wide range of biometric traits of several major categories (e.g. hard and soft biometrics, the static biometrics, the activity-related ones etc.), only specific modalities have been proven sufficient to support robust and accurate recognition performance up to now, i.e. fingerprint-, palmprint-, iris, face- and to an extent gait recognition.

1.1 Current Approaches

From the above, only face and gait recognition methods can be claimed to be less obtrusive since the subjects do not either directly interact with the recording sensor nor do they come in contact with it. Moreover, face recognition, that has long been and still one of the most active research areas in the recent years exhibits much higher recognition results than gait.

In this extend, face recognition has long been and still is one of the most active research areas in the recent years. Performance of $2D$ face matching systems depends on their capability of being insensitive to critical factors such as facial expressions, makeup, and aging, but mainly hinges upon extrinsic factors such as illumination differences, camera viewpoint, and scene geometry [Zhao2003]. However, provided the inherent limitations of $2D$ face matching, many researchers have stood for more environmental invariant recognition approaches. As such, the exploitation of the geometry of the anatomical structure of the face rather than its projective appearance has been a growing field of research recently [Smeets2010], via the implementation of efficient $3D$ transformation techniques [Mian2007] and corresponding face matching algorithms and systems [Berretti2010].

In order to improve the performance of typical biometric systems several methods have been proposed in the literature suggesting either multi-modal fusion [Ross2003b], generation of the user-specific fusion factors [Aggarwal2008], or seamless combination of heterogenous characteristic feature (e.g. anthropometric characteristics) in bayesian inference frameworks [Drosou2012].

1.2 Motivation

The motivation behind the current work is the need for robust and efficient face recognition systems that can address everyday security applications (e.g. PC login, gaming, etc.) with low-cost cameras. Most of the proposed approached have been evaluated with high resolution images or samples dense $3D$ point clouds, thus making it difficult to verify their validity and applicability in regular real world conditions, where the recorded images are not only captured in low resolution but also in noisy environments.

In this respect, the current paper proposes a thorough preprocessing of the recorded images, so as to improve their quality via efficient denoising and trimming techniques towards increasing their recognition capacity. Moreover, existing geometric feature extraction techniques effectively fused, so as to deliver multi-feature facial recognition with augmented performance under demanding scenarios of real world applications.

2 System Overview

An overview of the structure of the paper is illustrated in Figure 1 [1]. Initially, the raw $3D$ facial information is recorded and iteratively processed, so as to restore the face in frontal view and to deliver a smooth facial surface with no holes in it. Then, two state of the art algorithms (i.e. the $3DWW$ transform and the Spherical Face Representation) are

[1] Due to the restricted length of the current paper, the Figures referenced herein can be found under the following repository http://www.iti.gr/~drosou/RobustFaceRecLR

utilized for extracting discriminative geometrical features, while the produced biometric signature is enhanced via the generation of the corresponding cohort coefficients. Finally, the recognition decision is based on the comparison of this signature with the gallery template that refers to the claimed ID.

The current paper is organized as follows. A detailed description of the pre-processing algorithms for the generation and smoothing of the reconstructed facial surface is included in Section 3, while the core processing algorithm dealing with the extraction of the facial geometric characteristics is discussed in Sections 4.1 and 4.1, correspondingly, so as to make the paper self-consistent. Next, the estimation of the supplementary cohort coefficients follows in Section 4.3. A short description of the experiment and the utilized database follows in Section 5.1, and the experimental results and the contribution to the recognition performance of the combined approach proposed is exhibited in Section 5.2. Finally, the conclusions are drawn in Section 6.

3 Preprocessing

Initially, a point cloud of the whole captured setting is generated from depth and colour images provided by the Microsoft Kinect Sensor. The facial point cloud that contains useful information for user recognition is segmented from the rest of the image (i.e. noisy information from the facial images, such as areas with hairs or background areas), by preserving these $3D$ points that are included within the sphere with a radius of $\sim 10cm$ that is centered on the location of the nosetip (see Section 3.1).

This way, the work of Berretti et al. [Berretti2010] has been enriched with a preprocessing step for drawing face-specific ellipses, as it can be shown in the $1^{st} column$ of Figure 2(a). Once the facial region is segmented from the background, the triangulation of the remaining point cloud follows.

Yet, before extracting the geometric facial features that will be used as biometric descriptors, some further preprocessing is required. In particular, due to possible occlusions during the image capturing or due to noise induced by the materials of the sensor and the environmental context, some missing facial information may occur. In order to compensate this, a moving average window of 5×5 pixels is iteratively applied on the surface until all gaps are filled (Figure 2(b)).

Then, the rotation of the facial surface follows via the PCA algorithm on the $3D$ points, followed by the application of a uniform resampling algorithm. This procedure is iteratively applied until convergence (i.e. no further rotation of the point cloud occurs), as shown in Figure 2(c). The original face $3D$ points are placed in a perspective field due to the camera lens distortion. This results into facial point clouds with different resolutions. Via uniform sampling it is ensured that the faces have the same resolution before they compared. Differences in resolution of the faces can bias the similarity scores in favour of faces that are more densely sampled. The uniform sampling uses a 2d grid on the x and y planes, in which each cell is placed $1mm$ apart from its neighbours. In our experiments the average gap distance before the point cloud was about $3mm$, thus the uniform sampling meant also an oversampling operation.

Next, the resulted surface undergoes a final smoothing step via a moving median window with a size 3×3 pixels. The finally trimmed facial surface f is shown in Figure 2(e)

3.1 Nose tip detection

The step for the detection of the nose tip precedes the background segmentation and is initially based on an initial detection of the location of the nose tip (N'_{kinect}) from the coloured image, as delivered by the Kinect SDK toolkit. However, since the detection accuracy of this algorithm is not sufficient (red spots in Figure 3) a post processing algorithm has been initiated. I particular, all points within a sphere of $4cm$ around N'_{kinect} are undergone a PCA transformation and the new nose tip location is calculated as the median value of the M closest points to the origin of the depth axis. Then, by mapping this depth value on the initial surface, one can easily estimate a good approximation of the actual nose tip location ($N(x_0, y_0, z_0)$), as indicated by the blue spots in Figure 3.

4 Geometric Features Extraction

The $3D$ geometrical characteristics of the face of the user are extracted according to the algorithms presented in [Berretti2010] and in [Mian2007], respectively. In order to deliver a self consistent paper, a short description of these algorithms is included hereafter.

4.1 Intrafacial Directional Indices

The intrafacial Directional Indices of a $3D$ surface are extracted by estimating the $3D$ Walking Walkthroughs ($3DWW$) on it, as described below. Initially, the shortest geodesic distances of each point on the facial surface f with respect to the detected nosetip location $N(x_0, y_0, z_0)$ is estimated via the Dijkstra algorithm. This way, isogeodesic stripes of equal width (i.e. $1cm$) are formed, concentric and centered on the nose tip (1^{st} row in Figure 4). Thereafter, the so-called $3DWW$ are computed between all pairs of isogeodesic stripes (interstripe $3DWW$) and between each stripe and itself (intrastripe $3DWW$), as described in [Berretti2010]. In particular, the $3DWWs$ are computed as aggregate measures (i.e. Directional Indices) that provide a representation for the mutual displacement between the set of points of two spatial entities (i.e. isogeodesic stripes). Finally, these Directional Indices are cast to a graph representation x_{3DWW}, where stripes are used to label the graph nodes and $3DWWs$ to label the graph edges.

This way, the face recognition problem is reduced to a graph matching issue, suitable for very large data sets. Thereby, the similarity score $S(x_{3DWW}, \omega)$ between two face-related graphs is the combination of both the inter- and intra-stripe $3DWWs$ similarity measures.

4.2 Spherical Face Representation

The Spherical Face Representation (SFR) [Mian2007] is an integral non-invertible transform that can be seen as an extension of Circular Integration Transformation (CIT) in the $3D$ space. The SFR is used due to its aptitude to represent meaningful shape characteristics. The location of the nose tip (x_0, y_0, z_0) is detected following the approach in Paragraph 3.1 and is used as the center of integration for the utilized transformation method as shown in the 2^{nd} row of Figure 4.

In particular the $3D$ vector representing the facial surface f is transformed as shown by the following equation to an $1D$ vector, each element of which represents the number of pixels within the boundaries defined by two successive spheres (i.e. “rings”) with radius

$k\Delta\rho$ and $(k+1)\Delta\rho$, respectively.

$$x_{SFR} = SFR(\Delta\rho, t_1, t_2) = \frac{1}{\mathrm{T}_1}\frac{1}{\mathrm{T}_2}\sum_{k=1}^{K}\sum_{t_1=1}^{\mathrm{T}_1}\sum_{t_2=1}^{\mathrm{T}_2}$$

$$f(x_0 + k\Delta\rho\cos(t_1\Delta\theta)\sin(t_2\Delta\varphi), y_0 + k\Delta\rho\sin(t_1\Delta\theta)\sin(t_2\Delta\varphi), z_0 + k\Delta\rho\cos(t_2\Delta\varphi)) \tag{1}$$

for $k = 1, ..., K$ with $T_1 = 360^o/\Delta\theta$ with $T_2 = 360^o/\Delta\phi$, where $\Delta\rho$, $\Delta\theta$ and $\Delta\phi$ are the constant step sizes of the radius and angles variables and finally $K\Delta\rho$ is the radius of the smallest sphere that encloses the facial surface f.

The similarity measure between two facial surfaces is computed as the $L1$-distance score $S(x_{SFR}, \omega)$ between the current signature x_{SFR} and the template of the claimed ID ω.

4.3 Cohort

An important issue that may lower the performance of fusion based approaches deals with the biometric classes that are not compact with respect to the inter-class distances and not similarly distributed. In particular, when their distributions vary across identities, the recognition threshold may become too stringent for a few classes or too lenient for others. Moreover, their anisotropic distribution around the available samples renders it difficult to set a robust threshold separately for each class.

In this respect, the cohort biometric templates (i.e. neighboring signatures in terms of high similarity factor) have been suggested in [Aggarwal2008], so as to initiate an efficient fusion approach via the definition of user-specific scaling factors that are based on the inter-similarity scores of a genuine signature of the user with the most similar impostor signatures of a training dataset.

Thus, having knowledge of the cohort of each enrolled identity, the similarity of a query with the claimed identity is computed as the ratio of its raw similarity with the claimed identity divided by the raw similarity with the cohort of the claimed identity ω

$$S(x, \omega) = \frac{s(x, \omega)}{s(x, \bar{\omega})} \tag{2}$$

where $s(x, \bar{\omega})$ is the similarity score of the query with the cohort. The raw similarity with the claimed identity can directly be determined using the available matcher. Assuming the cohort set to be of size k, $s(x, \bar{\omega})$ is determined using the following max-rule

$$s(x, \bar{\omega}) = \max\{s(x, \omega^1), s(x, \omega^2), ..., s(x, \omega^k)\} \tag{3}$$

where $s(x, \omega^1), s(x, \omega^2), ..., s(x, \omega^k)$ is the set of similarity scores of the query with the cohort for the enrolled identity ω.

Herein, the two types of utilized face-related biometric features (i.e. $3DWW$ and SFR) are treated independently for cohort normalization and 5 cohort signatures are utilized for each separate biometric feature. Finally, the combined score is obtained by fusing the final cohort normalized scores of individual biometrics according to the late fusion technique proposed in [Aggarwal2008], where separate cohort sets for each biometric can be independently selected.

$$S_f(x,\omega) = f(S_{3DWW}(x,\omega), S_{SFR}(x,\omega)) \tag{4}$$

where $S_f(x,\omega)$ denotes the final combined score of the two biometrics and f is a fusion function like simple sum rule or product rule. $S_{3DWW}(x,\omega)$ and $S_{SFR}(x,\omega)$ denotes the cohort normalized scores of the two individual biometrics as determined using Eq.(2).

5 Experimental Results

The the selected recognition protocols along with the utilized dataset are described below (Section 5.1), while the performance of the system as a whole, as well as its performance when enabling each geometric feature separately are discussed in Section 5.2.

5.1 Dataset

The evaluation presented in this paragraph refers to the first session of the proprietary dataset "BIOTAFTOTITA". This database was captured in an indoor environment and includes various poses, angles (e.g. $-90^o, 45^o$), and grimaces (e.g. neutral, smile, scream, etc.) of the $3D$ recorded faces, under different lightning conditions, for both enrollment ("gallery") and authentication ("probe") procedures. Moreover, the first recorded session of the database consists includes 54 subjects. All $3D$ related recordings have been exclusively performed via the Microsoft Kinect Sensor®. Although the utilized $3D$ face matching algorithm exhibits high robustness in difficult environmental conditions and strange poses, herein, frames with neutral poses in 0^o have been selected for both gallery and probe for the initial evaluation of the proposed algorithm. In particular, for each scenario (e.g. neutral, smile, scream, etc.) multiple sessions have been recorded. Only one of the "neutral" sessions is selected to be used as the gallery and specifically only the five most discriminative frames of it. The distinctiveness is evaluated by creating a confusion matrix with the similarity measures between all frames. This way, 5 frames of this session are selected to be included in the biometric signature, while all other sessions and scenarios are used only for testing.

5.2 Results

In the current session the behaviour of the proposed system is presented as it is evaluated under a series of different scenarios. For reasons of brevity, from now on the term "gallery" will refer to the set of reference recorded images, whereas the term "probe" will stand for the test frames to be verified or identified.

It should be noted that different settings (e.g. stripes number for $3DWW$ estimation or k-step for the SFR algorithm, etc.) have been used when the system was functioning for identification purposes (i.e. Cumulative Match Characteristic (CMC) Curves) than when it was functioning in authentication mode (i.e. Receiver Operating Characteristic (ROC) Curves[2]).

In order to exhibit the contribution of the proposed preprocessing steps in the recognition performance of the current biometric system, the following experiments have been conducted. Specifically, the feature extraction algorithm and the matching process have been

[2] In the current paper, a ROC Curve is plotted, without loss of generality, as the function of the False Rejection Rate (y-axis) to the False Acceptance Rate (x-axis).

applied on (i) the raw reconstructed face as depicted in Figure 2(a), (ii) the reconstructed $3D$ face after the gaps have been filled as depicted in Figure 2(b), (iii) the reconstructed $3D$ face after face alignment, uniform sampling, gap filling and after the first iteration of the rotating and uniform resampling algorithm and (iv) the fully rotated reconstructed face (Figure 2(d)) without the final smoothing step.

The improvements in the performance of the two state-of-the-art algorithms presented in [Mian2007] and [Berretti2010], as well as the proposed combined approach are presented in terms of authentication and identification capacity in Figure 5(a), Figure 5(b), Figure 5(c) and Figure 6(a), Figure 6(b), Figure 6(c), respectively.

Some slight advances of the no-preprocessing performance over the after-gap-filling preprocessing performance can be explained by the fact, that in the neutral face in frontal-view recording protocol, that is examined herein, the gaps on the reconstructed face of the gallery recordings coincide with the ones in the probe recordings due to similar view angle and do not reflect the general case.

The most common scenario for face recognition refers to the capturing of images depicting the neutral (i.e. no grimace) face of a user in frontal view. Provided that the gallery images have been captured under the same conditions and protocol, it is expected that the performance of the system will be at its maximum. Indeed, as the reader can notice in Figure 7(a) and in Figure 7(b) the identification and the authentication rates are 100% and 0.25%, respectively. As expected, the combination of both types of the aforementioned geometric features (see Section 4) improves both the authentication and identification performance of the system.

The corresponding improvement in the performance of the combined system can be noted in the distributions of the matching scores of the clients (i.e. blue coloured bars) and the corresponding ones (i.e. red coloured bars) of the impostors, as shown in Figure 8(c). The corresponding distribution when each feature is utilized independently are shown in Figure 8(a) and Figure 8(b), for the Directional Indices and the SFR transformation, respectively.

The evaluation of the recognition performance of the proposed system in some more difficult scenarios for user is presented hereafter. At this point, it should be noted that the gallery images are always the same and have been recorded according to the neutral face in frontal view protocol, described above.

In this respect, Figure 9(a) and Figure 9(b) present the identification and authentication performance of the system in two more difficult scenarios. Specifically, one protocol indicates that the users should wear glasses when they undergo a recognition process, while the other one indicates a bad illumination in the environment when the recording is performed.

Two protocols that address cases of facial deformation with respect to the enrollment recordings indicate the yawning and the smiling of the user when his face is recorded in frontal view. The identification and authentication performance of the system in this cases are illustrated in Figure 10(a) and Figure 10(b), respectively.

Finally, a very demanding protocol is the one that deals with the face rotated with respect to the frontal view. The recognition potential falls even more when a grimace is performed by the users in parallel with the rotation of their heads. The identification and authentication performance of the system under these scenarios is analytically illustrated in Figure 11(a) and Figure 11(b), respectively.

6 Conclusion

An efficient and fast (real-time) methodology for robust user recognition based on two different types of geometric features related to the facial surface of the users has been proposed. Hereby, two fast transformations (i.e. $3DWW$ and SFR) of the $3D$ facial surface were utilized are seamlessly combined along with the corresponding cohort templates, so that the recognition performance of the system exhibits high potential even in the most difficult scenarios. The most important contribution of the current work refers to the preprocessing of the extremely low resolution facial image (i.e. $\sim 120 \times 90$ pixels per frame), so as to produce a smooth and trimmed continuous $3D$ facial surface. Future work of the current paper include the application of the proposed algorithm on the full version of the "BIOTAFTOTITA" database (Session 1 & 2), which includes 80 different subjects in total. Moreover, the system will be benchmarked in larger databases, so as to verify its robustness for real-world applications.

Acknowledgment

This work was supported by the Greek, nationally funded, research project "BIOTAFTOTITA".

References

[Ross2003a] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115-2125, 2003.

[Tsalakanidou07] F. Tsalakanidou, S. Malassiotis, and M. G. Strintzis, "A 3D face and hand biometric system for robust user-friendly authentication," *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2238–2249, 2007.

[Smeets2010] D. Smeets, P. Claes, D. Vandermeulen, J.G. Clement "Objective 3D face recognition: Evolution, approaches and challenges," *Forensic Science Intern.*, vol. 201, pp. 125–132, 2010.

[Zhao2003] W. Zhao, R. Chellappa, P.J. Philips, and A. Rosenfeld "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, no. 4 pp. 399–468, 2003.

[Ross2003b] A. Jain, K. Nandakumara, A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270-2285, 2005.

[Drosou2012] A. Drosou; N. Porfyriou; D. Tzovaras; , "Enhancing 3D face recognition using soft biometrics," *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp.1-4, 2012.

[Berretti2010] S. Berretti, A. D. Bimbo, I. C. Society, P. Pala, "3D Face Recognition Using Isogeodesic Stripes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 12, pp. 2162–2177, 2010.

[Mian2007] A. S. Mian, M. Bennamoun, R. Owens, "An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 11, pp. 1927–1943, 2007.

[Aggarwal2008] G. Aggarwal, N. K. Ratha, R. M. Bolle; R. Chellappa, "Multi-biometric cohort analysis for biometric fusion," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5224 – 5227, 2008.